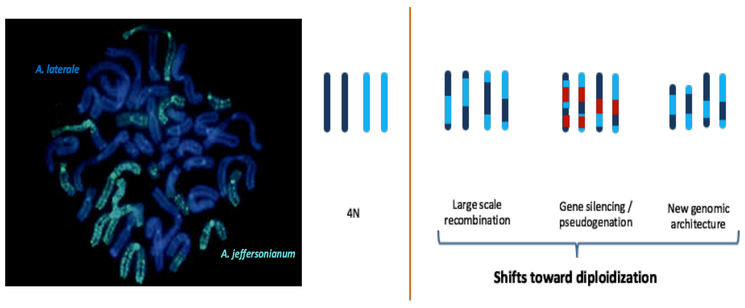
de novo assembly and annotation of *Ambystoma laterale* and *Ambystoma jeffersonianum* transcriptomes: The first steps toward investigating polyploid salamander expression

S.A. Sanders, M.E. Pfrender

Abstract

Polyploid Ambystoma salamanders are interesting as stable polyploid vertebrate system. Polyploidy in vertebrates is not well understood, and the unique salamander complex features non-recombining genomes with varying levels of ploidy and contribution of parental species. We assembled and annotated the two major contributors to this complex, *Ambystoma jeffersonianum* and *A. laterale* in preparation for investigating how the genomes respond to diverse polyploidy. We assembled each transcriptome using four assemblers (Velvet, SOAPdeNOVO, Trinity, and TransAbyss). We then curated and annotated both transcriptomes using tr2aacds (EviGenes) and KEGG. Orthologs and differentially expressed genes between the two parental species were identified to establish a baseline by which to compare the polyploids. Additionally, we developed an immune inventory for the two transcriptomes. The annotation will be useful in the next steps as we investigate the polyploid salamander libraries to determine the effect of polyploidy on expression.



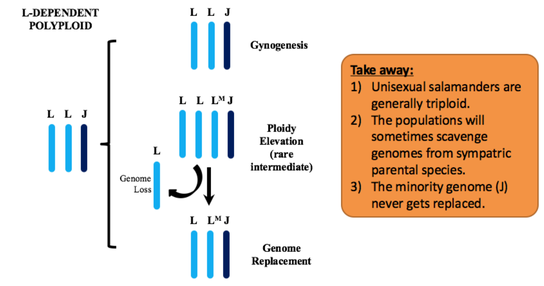
Left, GISH chromosomes of a LLJ polyploid salamander, with blue chromosomes being Lat and green being Jef (courtesy of J. Bogart). Right, chromosome architecture of polyploid salamanders versus changes that occur as polyploids diploidize.

◄ Why polyploids?

Polyploidy is a major driver of evolution (Comai 2005), but the dynamics of early stages of interaction are not well understood, especially in vertebrate polyploids. We seek to understanding how complete genome networks function, interact, and conflict in normal and stress conditions before diploidization occurs.

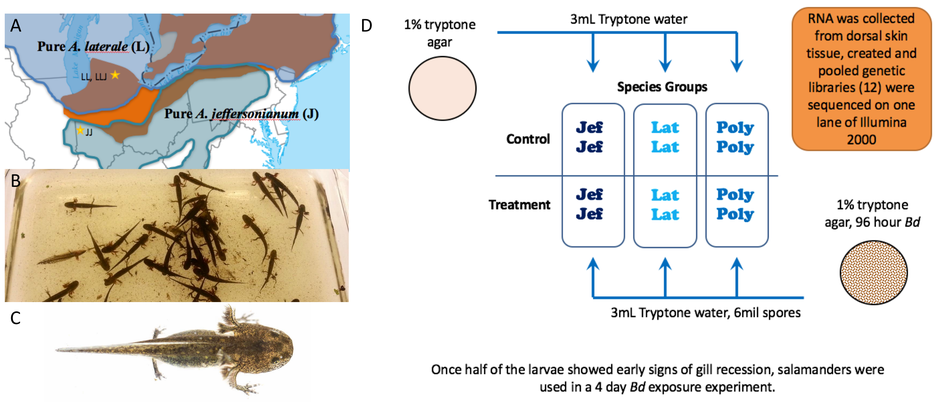
Why these polyploids? ►

Ambystoma polyploids are a stable polyploid system, with 3-5N individuals. Typically individuals are 3N, with two copies of one genome and one of non-local species. These genomes undergo very little recombination and are stable for many generations before one of the majority genomes (in this case L) is swapped out through semi-sexual reproduction. This results in a persistent case of complete genome networks coexisting.



Experimental design and Transcriptome Assembly

Ambystoma laterale (Lat) and *A. jeffersonianum* (Jef) were collected from breeding ponds (see map, orange is the polyploid range). The Lat and polyploid (poly) salamanders came from the same populations, the Jef came from one of the closest populations to the Lat population. Animals were all exposed to commercial antimicrobial treatment for two days before chytrid exposure began. Larvae were exposed to chytrid for three days (with water changes every 24 hours). RNA libraries were then created using standard methods. Stress was used as a means of expanding transcripts captured and to serve our future interest in the polyploid response to stress (chytrid zoospore exposure). Assembly of the parental species transcriptomes is a critical step before we can analyze the polyploid salamander libraries.



A) Map of sampling, where orange indicates polyploid range. B) Lat and Poly larvae from pond traps. C) Jef larvae. D) Experimental Design. Poly libraries were not analyzed as of yet. All samples from each Lat and Jef were used in assembly of transcriptomes.

Assembly

- Trinity 2.4.0, Velvet 1.2.10/ Oases 0.2.08, TransAbyss 1.5.5, and SOAP denovo 1.03
- Used k=35,45,55,65,75,85 kmers for all but Trinity (k=25 default)
- Combined all kmers and assemblers using EviGenes Pipeline (D. Gilbert, 2013).
- Kept all longest, unique, complete contigs ("okay complete") from EviGenes output.

Final Assembly Stats:

Lat: 28,792 contigs
- GC: 46.75%
- N50:3,540bp

Jef :28,617 contigs
- GC: 47.01%
- N50:3,599bp

Correctness

Correctness was measured by using blast similarity to published Axolotl and Newt transcriptomes (Abdullayev et al. 2013, Bryant et al. 2017).

Lat compared to Axolotl	
Matches	25,085
Avg Identity	93.80%
Avg Length	1035.57bp

Lat compared to Newt	
Matches	8,420
Avg Identity	83.88%%
Avg Length	1508.78bp

Jef compared to Axolotl	
Matches	25,003
Avg Identity	93.80%
Avg Length	990.39bp

Jef compared to Newt	
Matches	8,351
Avg Identity	83.98%
Avg Length	1514.65bp

Completeness

BUSCO v1.22 was used to determine number of single copy genes successfully recovered, compared to vertebrates and metazoans.

Lat	Vertebrate	Metazoan
Complete	67.52	85.41
Duplicate	1.52	7.23
Fragment	2.61	1.54
Missing	29.87	13.05

Jef	Vertebrate	Metazoan
Complete	66.99	84.58
Duplicate	1.46	8.19
Fragment	2.41	1.42
Missing	30.60	14.00

The relatively low duplicate and fragmentation is a good indicator of complete assembly. Missing genes are likely a result of single tissue use.

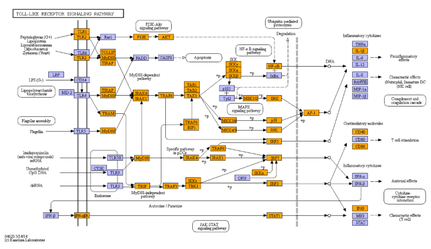
Analysis of Transcriptomes

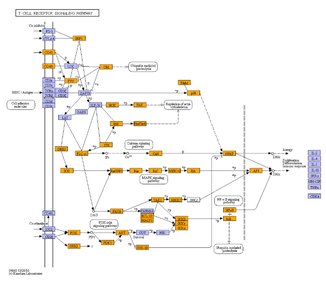
Annotation

Only longest unique complete ("okay complete") contigs were used in downstream analysis. The okay complete contigs for each species were submitted online to Ghost KOALA for KEGG annotation. Trinotate annotation is in progress and will add GO terms and PFAM groups to these annotations.

Ghost KOALA assigned 8,686 Lat and 8,499 Jef transcripts KEGG identifiers.

▼KEGG BRITE pathways showing the RBH orthologs in two major immune pathways of interest in chytrid response, Toll like receptors and T-cell mediated response (see Next Steps Figure). Orange is an identified RBH, blue is currently missing from RBH set.





Orthology

Ortholog identification was done using a **Reciprocal Best Hit (RBH)** pipeline leveraging Blast 2.5.0. Trinotate annotation will add annotation and increase ability to cross reference annotations in the future.

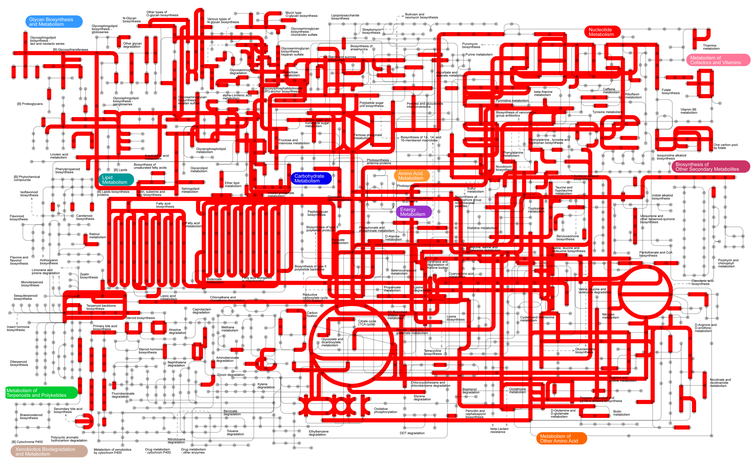
Reciprocal Best Hit Orthologs
- 11,339 ortholog pairs identified
- Average of 95.4138% identity over 1,037.83 bp
- 5,372 have KEGG identifiers (1,889 mapped below)
- 305 Lat and 299 Jef orthologs were identified as immune genes via KEGG.

Differential Expression (DE)

Differential Expression EdgeR using Trinity 2.4.0 package, a p-value cut off of 1e-3, and TPM as a count method. Control vs. Treatment analysis used all final transcripts; Lat vs. Jef analysis used only RBH orthologs only.

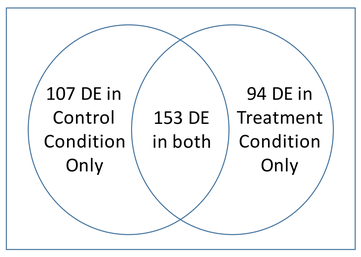
Control vs. Treatment Results:
- 35 Lat transcripts DE, 25 were KEGG immune
- Jef – 1 transcript DE, not immune

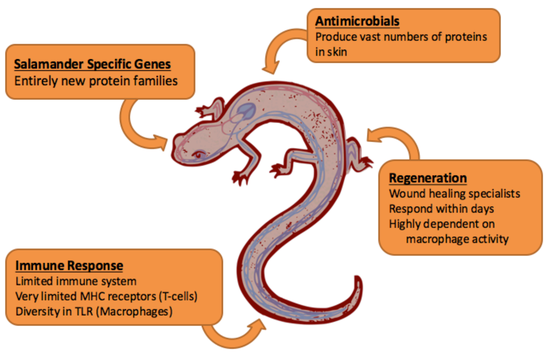
Lat vs. Jef Results:
- Control – 260 orthologs DE between Lat and Jef, 5 Immune
- Treatment – 247 orthologs DE between Lat and Jef, 4 immune
- 153 transcripts were DE in both conditions
- 94 transcripts were only DE in response to treatment (below), indicating differential response to the stressor.



◄KEGG map of all RBH orthologs identified using Ghost KOALA and IPATHv2, where red indicates presence.

▼Transcripts differentially expressed between Lat and Jef in control and treatment conditions.





Conclusions

1) Assemblies have limited fragmentation and duplication, cover a large portion of single copy genes as well as homologs to similar species – preliminary indications of good assemblies.
2) The 11,339 orthologous pairs identified will permit comparisons of expression in homeologs in polyploids. These orthologs span many pathways, expanding their utility and interest.
3) Differential response to conditions indicates target genes to investigate in polyploids – i.e. how do polyploids resolve different regulation of similar gene networks within the same cell?

◄ Next Steps

Assessment of polyploid libraries will begin after completion of annotation of parental genomes. We are interested in how Ambystoma salamanders in resist chytrid (see possible traits of interest to the left) and how these responses are skewed in polyploids.

Acknowledgements

Work was supported by the Center for Aquatic Conservation at University of Notre Dame and the National Center for Genomic Analysis Support at Indiana University, which is supported by the National Science Foundation under grant No. ABI-1458641 to Indiana University.

References

Adbdullayev et al. 2013. A reference transcriptome and inferred proteome for the Salamander Notophthalamus viridescens. Exp. Cell Res 319(8): 1187-1197.

Bryant et al. 2017. A tissue-mapped Axolotl de novo transcriptome enables identification of limb regeneration factors. Cell Reports. 18(3): 762-776.

Comai, L. 2005. The advantages and disadvantages of being polyploid. Nature Reviews Genetics. 6(11): 836--846.

EvidentialGene, Gilbert, D. tr2aacds, mRNA transcript assembly software. 2013.